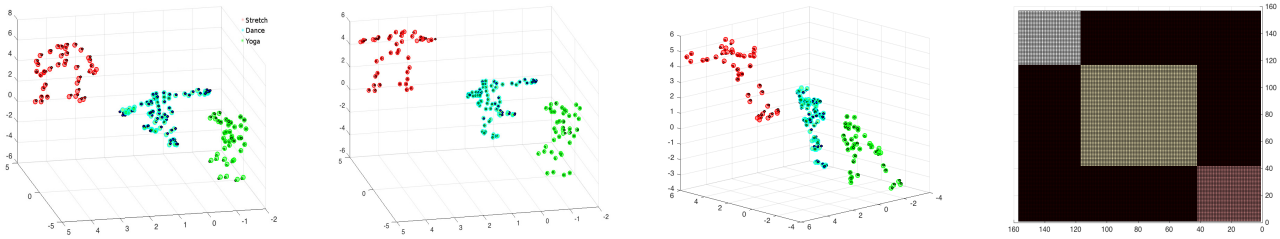
Multi-body Non-rigid Structure-from-Motion

Suryansh Kumar[1] Yuchao Dai[1] Hongdong Li[1,2]
[1]Research School of Engineering, Australian National University.
[2] Australian Centre for Robotic Vision.



(a) (b) (c) (d)

Figure 1: A nutshell of our framework. (a)-(c) Multi-body non-rigid structure from motion on the synthetic CMU MoCap dataset [2]. Our approach is able to reconstruct and segment each action such as stretch(red), dance(cyan) and yoga(green) faithfully with 3D reconstruction error of 0.0413 and 0 segmentation error. Here, different color corresponds to different segmentation, while dark and light color circles show ground-truth and reconstructed 3D coordinates respectively. (d) Block diagonal matrix obtained by using our approach.

Abstract

In this paper, we present the first multi-body non-rigid structure-from-motion (SFM) method, which simultaneously reconstructs and segments multiple objects that are undergoing non-rigid deformation over time. Under our formulation, 3D trajectories for each non-rigid object can be well approximated with a sparse affine combination of other 3D trajectories from the same object. The resultant optimization is solved by the alternating direction method of multipliers (ADMM). We demonstrate the efficacy of the proposed method through extensive experiments on both synthetic and real data sequences. Our method outperforms other alternative methods, such as first clustering the 2D feature tracks to groups and then doing non-rigid reconstruction in each group or first conducting 3D reconstruction by using single subspace assumption and then clustering the 3D trajectories into groups.

1. Introduction

Structure-from-Motion (SFM) targets at recovering 3D structure and camera motion from monocular 2D feature tracks. Conventional SFM primarily concerns with the 3D reconstruction of a single rigidly moving object seen by a static camera, or a static and rigid scene observed by a moving camera –in both cases there are only one relative rigid motion involved. Recent progress have extended rigid motion SFM to the areas of *multi-body* SFM [5][22] and significant improvement have been made over *single body* non-rigid SFM [3][21][2][6][9][4] (see Table.1). Along this line of thinking, there is apparently a missing gap of "multi-body non-rigid SFM". Hence, this paper proposes for the first time, an effective framework for multi-body NRSFM.

Most of the existing methods for NRSFM have implicitly assumed that there is only one deformable shape or object. However, real world scenarios are much more complicated involving multiple, independently deforming objects in the scene. Multiple nonrigid objects are commonly encountered in our daily lives, for example, in motion capture, multiple persons perform different activities with possible interactions (see Fig. 1 for example); in human-computer interaction, different users may conduct different gesture commands; in traffic scene, multiple vehicles and walking pedestrians create **multi-body non-rigid deformations**.

To handle such multiple non-rigid deformations in 3D reconstruction, a natural idea would be to simply represent the multiple non-rigid deformations as a single (though more complex) non-rigid deformation (with higher order or higher rank), and then apply any state-of-the-art non-rigid structure-from-motion methods such as [6][13]. However, by this idea, the inherent structure of the problem has not

been exploited, which may hinder the success of 3D reconstruction. Even if the method succeeds in obtaining 3D reconstruction, it cannot tell meaningful segmentation of multiple non-rigid objects. Another choice would be to conduct non-rigid motion segmentation [7] and non-rigid 3D reconstruction [6] successively. In this way, the solution of each sub-task does not benefit from the solution of the other subtask. Therefore, we would like to emphasize that since non-rigid deformation originally occurs in 3D space, it's more intuitive to perform non-rigid motion segmentation and reconstruction simultaneously in 3D space than solving this problem using two step process.

This paper introduces an approach to perform non-rigid 3D reconstruction and motion segmentation simultaneously. Specifically, we represent multi-body non-rigid motion as a union of 3D trajectory sub-spaces[1]. By using the self-expressiveness model in representing multiple linear or affine subspace, where each 3D trajectory can be expressed with other trajectories in the same subspace only, enables us in compact representation of trajectories. In this way, we are able to exploit the inherent grouping structure in *3D trajectory space*. For dense non-rigid reconstruction, we could further enforce the spatial coherence constraint. By contrast to existing methods, this endows us the following benefits: (a) A compact representation for component non-rigid deformation in 3D trajectory space. (b) Joint reconstruction and motion segmentation of multiple deformable objects. (c) Improved spatial regularity in 3D non-rigid dense reconstruction (in contrast, a hard segmentation of the 2D tracks may result in discontinuity at the segmentation boundary).

Contributions: (1) This paper is first to model multiple non-rigid deformations as points in the union of multiple affine 3D trajectory subspace. This enables us to jointly solve the non-rigid reconstruction and non-rigid motion segmentation problems in 3D trajectory space; (2) Our formulation can handle both sparse and dense multi-body non-rigid reconstruction problems uniformly; (3) We propose an efficient optimization procedure based on ADMM method.

2. Related Work

Ever since the seminal work by Bregler *et al*. [3] modeling a non-rigid shape as lying in a "*shape space*" (a linear combination of basis shapes), considerable progress has been made in the area of non-rigid 3D reconstruction. In 2004, Xiao *et al*. [21] showed the inherent ambiguity in modeling non-rigid shape and proposed a remedy of "basis constraints" to derive a closed-form solution. In 2008, Akhter *et al*. [2] presented a dual approach by modeling 3D trajectories, *i.e*. "*trajectory space*". In 2009, Akhter *et al*. [1] proved that even there is an ambiguity in shape bases or trajectory bases, non-rigid shapes can still be solved uniquely without any ambiguity. In 2012, Dai *et al*. [6] proposed a "prior-free" method to recover camera motion and 3D non-rigid deformation by exploiting low rank constraint only. Besides shape basis model and trajectory basis model, the shape-trajectory approach [11] combines two models and formulates the problems as revealing trajectory of the shape basis coefficients. Besides linear combination model, Lee *et al*. [13] proposed a Procrustean Normal Distribution (PND) model, where 3D shapes are aligned and fit into a normal distribution. Simon *et al*. [18] exploited the Kronecker pattern in the shape-trajectory (spatial-temporal) priors. Zhu and Lucey [25] applied the convolutional sparse coding technique to NRSFM using point trajectories. However, the method requires to learn an over-complete basis of 3D trajectories, prior to performing 3D reconstruction.

Despite of the above success, NRSFM is still far behind its rigid counterpart. This is mainly due to the difficulty in modeling real world non-rigid deformation. Real world non-rigid reconstruction generally requires the ability to handle long-term, complex and dense non-rigid shape variations. Such complex and dense non-rigid motion not only increases the computational complexity but also adds difficulty in modeling various kinds of different motions. Zhu *et al*. [24] proposed to represent long-term complex non-rigid motion as lying in a union of shape sub-spaces rather than sum of sub-spaces. Cho *et al*. [4] represented complex shape variations probabilistically by a mixture of primitive shape variations.

By contrast to the above methods dealing with sparse NRSFM, dense NRSFM methods such as [16][9][10][17] aim at achieving 3D reconstruction for each pixel in the video sequence, where spatial constraint has been widely used to regularize the problem. Garg *et al*. [9] presented a variational formulation to dense non-rigid reconstruction by exploiting the spatial smoothness in 3D shapes, which in principle deals with single non-rigid deformation in contrast to our multiple non-rigid deformations. Fragkiadaki *et al*. [8] solved the problem in sequel, namely, video segmentation by multi-scale trajectory clustering, 2D trajectory completion, rotation estimation and 3D reconstruction. Recently, Yu *et al*. [23] bridges template based method and feature track based method by proposing a dense template based direct approach to deformable shape reconstruction from monocular sequences.

Russell *et al*. [17] proposed to simultaneously segment a complex dynamic scene containing a mixture of multiple objects into constituent objects and reconstruct a 3D

[1]Zhu *et al*. [24] used the union of sub-spaces representation (different non-rigid deformations lie in different sub-spaces), where the subspace is defined in shape space contrast to our trajectory space. As we will show later, this difference provides uniqueness of our formulation in dealing with multiple non-rigid deformation and in dealing with dense case.

Table 1: A classification of different SFM problems defined by the number of objects and the rigidity of each object. This paper aims to fill in the currently missing work of **Multi-body Non-rigid SFM** shown in blue.

	Single body	Multi-body
Rigid	Single-body Rigid SFM [19] $\mathtt{W}_{2\mathtt{F}\times\mathtt{P}} = \mathtt{R}_{2\mathtt{F}\times 3}\mathtt{S}_{3\times\mathtt{P}}, \text{rank}(\mathtt{S}) = 3$	Multi-body Rigid SFM [5][22] $\mathtt{W}_{2\mathtt{F}\times\mathtt{P}} = \mathtt{R}_{2\mathtt{F}\times 3\mathtt{F}}\mathtt{S}_{3\mathtt{F}\times\mathtt{P}}, \text{rank}(\mathtt{S}) = 3\mathtt{K}$
Non-rigid	Single-body Non-rigid SFM [3] $\mathtt{W}_{2\mathtt{F}\times\mathtt{P}} = \mathtt{R}_{2\mathtt{F}\times 3\mathtt{F}}\mathtt{S}_{3\mathtt{F}\times\mathtt{P}}, \text{rank}(\mathtt{S}) = 3\mathtt{K}$	Multi-body Non-Rigid SFM $\mathtt{W}_{2\mathtt{F}\times\mathtt{P}} = \mathtt{R}_{2\mathtt{F}\times 3\mathtt{F}}\mathtt{S}_{3\mathtt{F}\times\mathtt{P}}, \mathtt{S} = \mathtt{SC}$. i.e, 3D trajectory should lie in union of linear/affine subspace.

model of the scene by formulating the problem as hierarchical graph-cut based segmentation, where the whole scene is decomposed into background and foreground objects and the complex motion of non-rigid or articulated objects are modeled as a set of overlapping rigid parts. Recently, piecewise rigid approach [15] for complex motion using consecutive images tried to solve this task in 3 consecutive steps (motion segmentation, 3D reconstruction and relative scale). Our method differs from these methods in the following aspects: 1) We provide a compact representation to multiple non-rigid deformation problem; 2) We propose an efficient and elegant optimization based on ADMM; 3) Our method could deal with both sparse and dense scenarios.

3. Formulation

We seek to reconstruct 3D trajectories such that they satisfy the union of affine subspace constraint (*i.e.* 3D trajectories lie in a union of affine subspaces) and non-rigid shape constraints (low rank and spatial coherent).

Let us consider a monocular camera observing multiple non-rigid objects. We use the *orthographic camera* model and eliminate the translation component in camera motion [3]. The image measurement $\mathtt{w}_{ij} = [u_{ij}, v_{ij}]^T$ and 3D point $\mathtt{S}_{ij}$ on the non-rigid shape are related by the camera motion $\mathtt{R}_i$ as: $\mathbf{w}_{ij} = \mathtt{R}_i\mathtt{S}_{ij}$, where $\mathtt{R}_i \in \mathbb{R}^{2\times 3}$ denotes the first two rows of the i-th camera rotation. Under this representation, stacking all the $\mathtt{F}$ frames of measurements and all the $\mathtt{P}$ points in a matrix form will give us:

$$\mathtt{W} = \mathtt{RS}, \tag{1}$$

where $\mathtt{R} = \text{blkdiag}(\mathtt{R}_1, \cdots, \mathtt{R}_F) \in \mathbb{R}^{2\mathtt{F}\times 3\mathtt{F}}$ denotes the camera motion. NRSFM aims at recovering the *camera motion* $\mathtt{R}$ and 3D non-rigid reconstruction $\mathtt{S} \in \mathbb{R}^{3\mathtt{F}\times\mathtt{P}}$ from the 2D *measurement matrix* $\mathtt{W} \in \mathbb{R}^{2\mathtt{F}\times\mathtt{P}}$ such that $\mathtt{W} = \mathtt{RS}$.

3.1. Representing multi-body non-rigid structure as a union of affine subspace

We assume that multiple non-rigid structures that correspond to distinct motion lie in a union of affine subspace. Here, the underlying assumption is that the trajectories belonging to different non-rigid objects span a distinct affine subspace. Figure 2(b) clearly validates such assumption as there are only connections within clusters and no connections between clusters or block diagonal structure.

Now, consider each trajectory $\mathtt{S}_j$ that corresponds to a $3\mathtt{F}$ dimensional vector formed by stacking the 3D tracks of feature point j across all frames.

$$\mathtt{S}_j = [\mathtt{S}_{1j}^T, \mathtt{S}_{2j}^T, ...\mathtt{S}_{\mathtt{F}j}^T]^T \in \mathbb{R}^{3\mathtt{F}\times 1} \tag{2}$$

where $\mathtt{S}_{fj} \in \mathbb{R}^{3\times 1}$ with f varies from $\{1, 2, 3...\mathtt{F}\}$. Under the union of affine subspace representation for multi-body non-rigid reconstruction, feature trajectories associated with each non-rigid motion lie in an affine subspace. After taking $\mathtt{P}$ such $3\mathtt{F}$ dimensional trajectory and stacking into the column of a matrix we form $\mathtt{S}$ matrix $\in \mathbb{R}^{3\mathtt{F}\times\mathtt{P}}$. Mathematically, it implies that each column of $\mathtt{S}$ is drawn from a union of n subspace in $\mathbb{R}^{3\mathtt{F}}$. Therefore, each trajectory in a union of affine subspace can be faithfully reconstructed by combination of other trajectories in the same subspace. This leads to the $self\text{-}expressiveness$ of the 3D trajectories. Concretely,

$$\mathtt{S}_j = \mathtt{SC}_j, \mathtt{C}_{jj} = 0. \tag{3}$$

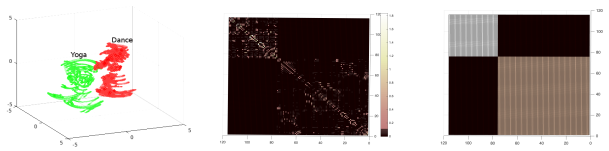
Here, $\mathtt{C}_j$ is $\mathtt{P}\times 1$ coefficient vector and $\mathtt{C}_{jj} = 0$ takes care of trivial solution. Stacking all such coefficient vectors we form a $\mathtt{C}$ matrix $\in \mathbb{R}^{\mathtt{P}\times\mathtt{P}}$ that captures the similarity between different trajectories. Using the fact that any trajectory of $\mathtt{S}$ in an affine subspace can be written as an affine combination of other trajectories from $\mathtt{S}$ and to cluster trajectories that lies near to union of affine subspace, we arrive at the following equation.

$$\mathtt{S} = \mathtt{SC}, 1^T\mathtt{C} = 1^T, \text{diag}(\mathtt{C}) = 0. \tag{4}$$

Figure 2 shows the affinity matrix $\mathtt{A} = |\mathtt{C}| + |\mathtt{C}^T|$ obtained for two objects that undergo non-rigid deformation. The solution clearly shows that multi-body non-rigid structure can be represented as union of affine subspace.

3.2. Representing multiple non-rigid deformations in case of sparse feature tracks

To solve the problem of multi-body NRSFM in case of sparse feature tracks, we propose the following optimiza-



(a) (b) (c)

Figure 2: (a) Two subjects performing different non-rigid motion that are dance and yoga. Red and green color shows the entire trajectory of each objects over F frames. (b) Visualization of affinity matrix obtained using our formulation. (c) Clean affinity matrix obtained after incorporating spectral clustering. [14]. Best viewed on screen.

tion framework for simultaneous reconstruction and segmentation of objects that are undergoing non-rigid deformation.

$$\begin{aligned} &\underset{\mathtt{C},\mathtt{S},\mathtt{S}^\sharp}{\text{minimize}} \frac{1}{2}\|\mathtt{W}-\mathtt{RS}\|_F^2+\lambda_1\|\mathtt{C}\|_1+\lambda_2\|\mathtt{S}^\sharp\|_* \\ &\text{subject to:} \\ &\mathtt{S}^\sharp=g(\mathtt{S}),\mathtt{S}=\mathtt{SC},1^T\mathtt{C}=1^T,\text{diag}(\mathtt{C})=0. \end{aligned} \tag{5}$$

The first term in the above optimization is meant for penalizing re-projection error under *orthographic* projection. Under single-body NRSFM configuration, 3D shape $\mathtt{S}$ can be well characterized as lying in a single low dimensional linear subspace. However, when there are multiple non-rigid objects, each non-rigid object could be characterized as lying in an affine subspace. One can argue that affine subspace of dimension n can be considered as a subset of $(n+1)$ dimension that includes origin. Nonetheless, such representation may result in ambiguous solution while clustering different subspace. The fact that the 3D trajectories lie in a union of affine sub-spaces as argued previously §3.1 we put the Eq. 4 as our optimization constraints.

In addition to this, to reveal the intrinsic structure of multi-body non-rigid structure-from-motion (NRSFM), we seek for the sparsest solution of $\mathtt{C}$ ([7]). So, the second term in **Eq.** (5) enforces l_1 norm minimization of $\mathtt{C}$ matrix. Lastly, we enforce a global shape constraint for compact representation of multi-body non-rigid objects by penalizing the rank of entire non-rigid shape. Similar to [6] and [9], we penalize the nuclear norm of the reshuffled shape matrix $\mathtt{S}^\sharp \in \mathbb{R}^{\mathtt{F}\times 3\mathtt{P}}$, this is because nuclear norm is known as the convex envelope of the rank function. Here, $\mathtt{g}(\mathtt{S})$ denotes the mapping from $\mathtt{S}\in\mathbb{R}^{3\mathtt{F}\times\mathtt{P}}$ to $\mathtt{S}^\sharp\in\mathbb{R}^{\mathtt{F}\times 3\mathtt{P}}$.

3.3. Representing multi-body non-rigid deformations in case of dense feature tracks

When per pixel feature tracks are available, we can enforce spatial regularization ($Markovian\ assumption$) i.e

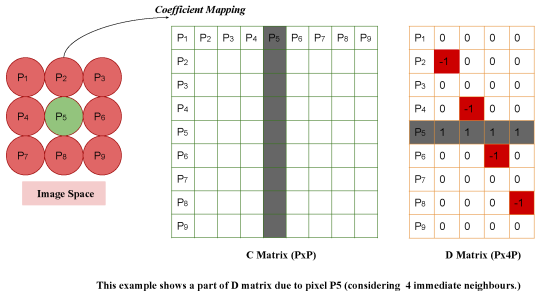


Figure 3: D matrix caters the neighboring trajectory relation. In the above illustration $\mathtt{P}_2$, $\mathtt{P}_4$, $\mathtt{P}_6$, $\mathtt{P}_8$, are the 4 immediate neighboring trajectories of $\mathtt{P}_5$. Therefore, corresponding elements of $\mathtt{D}$ matrix has -1 entries, $\mathtt{P}_5$ has value $\mathtt{1}$ and the rest entries are $\mathtt{0}$. Therefore, the corresponding coefficient column of $\mathtt{C}$ matrix now rely on relations defined in the $\mathtt{D}$ matrix. Here, $\mathtt{C}$ and $\mathtt{D}$ are $\mathtt{P}\times\mathtt{P}$ and $\mathtt{P}\times\mathtt{4P}$ matrix respectively. Where, $\mathtt{P}$ is the number of total feature tracks.

there will be a high correlation between neighboring features. To exploit this property, the spatial smoothness can be used as a regularization term to further constrain the non-rigid reconstruction. Garg *et al*. [9] proposed to use the total variation of the 3D shape $\|\mathtt{S}\|_{TV}$. By contrast, we propose to enforce the spatial smoothness constraint on the coefficient matrix $\mathtt{C}$ directly by using the L_1 norm,

$$\sum_{(i,j)\in\mathcal{N}} \|\mathtt{C}_i-\mathtt{C}_j\|_1, \tag{6}$$

i.e., the total variation of $\mathtt{C}$. This definition gives us the benefit in solving the problem as proved later. In essence, the total variations of $\mathtt{C}$ and $\mathtt{S}$ are correlated. However, it is desirable that $\mathtt{C}$ matrix must cater the self-expressiveness of the non-rigid shape deformation as compact as possible. So, we incorporate spatial smoothness constraint on coefficient matrix rather than on shape matrix $\mathtt{S}$.

By introducing an appropriately defined matrix $\mathtt{D}$ encoding the neighboring relation, Eq. (6) can be expressed as:

$$\|\mathtt{CD}\|_1=\sum_{(i,j)\in\mathcal{N}} \|\mathtt{C}_i-\mathtt{C}_j\|_1. \tag{7}$$

In Fig.3, we illustrate the process of how to obtain the matrix $\mathtt{D}$. By incorporating this spatial constraint to the optimization equation (5), that facilitates this neighboring constraint, we reach the following optimization for dense tracks:

$$\begin{aligned} &\underset{\mathtt{C},\mathtt{S},\mathtt{S}^\sharp}{\text{minimize}} \frac{1}{2}\|\mathtt{W}-\mathtt{RS}\|_F^2+\lambda_1\|\mathtt{C}\|_1+\lambda_2\|\mathtt{CD}\|_1+\lambda_3\|\mathtt{S}^\sharp\|_* \\ &\text{subject to:} \\ &\mathtt{S}^\sharp=g(\mathtt{S}),\mathtt{S}=\mathtt{SC},1^T\mathtt{C}=1^T,\text{diag}(\mathtt{C})=0. \end{aligned} \tag{8}$$

4. Solution

Due to the bilinear term $\mathtt{S} = \mathtt{SC}$, the overall optimization of Eq.-(8) is non-convex. We solve it via the ADMM, which has a proven effectiveness for many non-convex problems and is widely used in computer vision. The ADMM works by decomposing the original optimization problem into several sub-problems, where each sub-problem can be solved efficiently. To this end, we seek to decompose Eq.-(8) into several sub-problems.

First note that the two L_1 terms $\|\mathtt{C}\|_1$ and $\|\mathtt{CD}\|_1$ can be put together as $\|\mathtt{C}[\mathtt{I}\ \mathtt{D}]\|_1$. Without loss of generality, we still denote the new term as $|\mathtt{CD}|_1$, the only difference is, the new dimension of $\mathtt{D}$ will be $\mathtt{P} \times 5\mathtt{P}$, thus the cost function becomes: $\frac{1}{2}\|\mathtt{W} - \mathtt{RS}\|_F^2 + \lambda_1\|\mathtt{CD}\|_1 + \lambda_2\|\mathtt{S}^\sharp\|_*$. To further decouple the constraint, we introduce an auxiliary variable $\mathtt{E} = \mathtt{CD}$. With these operations, the optimization problem Eq.-(8) can be reformulated as:

$$\begin{aligned} &\underset{\mathtt{E},\mathtt{S},\mathtt{S}^\sharp,\mathtt{C}}{\text{minimize}}\ \frac{1}{2}\|\mathtt{W} - \mathtt{RS}\|_F^2 + \lambda_1\|\mathtt{E}\|_1 + \lambda_2\|\mathtt{S}^\sharp\|_* \\ &\text{subject to:} \\ &\mathtt{S}^\sharp = g(\mathtt{S}), \mathtt{S} = \mathtt{SC}, \mathtt{CD} = \mathtt{E}, 1^T\mathtt{C} = 1^T, \mathrm{diag}(\mathtt{C}) = 0. \end{aligned} \tag{9}$$

The Augmented Lagrangian formulation for Eq.-(9) is:

$$\begin{aligned} \mathcal{L}(\mathtt{S},\mathtt{S}^\sharp,\mathtt{C},\mathtt{E},\{\mathtt{Y}_i\}_{i=1}^4) = \frac{1}{2}\|\mathtt{W} - \mathtt{RS}\|_F^2 + \lambda_1\|\mathtt{E}\|_1 + \\ \lambda_2\|\mathtt{S}^\sharp\|_* + <\mathtt{Y}_1, \mathtt{S}^\sharp - g(\mathtt{S})> + \frac{\beta}{2}\|\mathtt{S}^\sharp - g(\mathtt{S})\|_F^2 + \\ <\mathtt{Y}_2, \mathtt{S} - \mathtt{SC}> + \frac{\beta}{2}\|\mathtt{S} - \mathtt{SC}\|_F^2 + <\mathtt{Y}_3, \mathtt{CD} - \mathtt{E}> + \\ \frac{\beta}{2}\|\mathtt{CD} - \mathtt{E}\|_F^2 + <\mathtt{Y}_4, 1^T\mathtt{C} - 1^T> + \frac{\beta}{2}\|1^T\mathtt{C} - 1^T\|_F^2. \end{aligned}$$

where $\{\mathtt{Y}_i\}_{i=1}^4$ are the matrices of Lagrange multipliers corresponding to the four equality constraints, and β is a penalty parameter. We do not need to introduce a Lagrange multiplier for the diagonal constraint of $\mathrm{diag}(\mathtt{C}) = 0$ as we will enforce this constraint exactly in the solution of $\mathtt{C}$.

The ADMM iteratively updates the individual variable so as to minimize $\mathcal{L}$ while the other variables are fixed. In our formulation variables $\mathtt{S}$, $\mathtt{S}^\sharp$, $\mathtt{E}$, $\mathtt{C}$ are solved using the following sub-problems. Kindly, refer to supplementary material for detailed solution.

$$\mathtt{S}_{i+1} = \arg\min_{\mathtt{S}} \mathcal{L}(\mathtt{S}_i, \mathtt{S}^\sharp, \mathtt{C}, \mathtt{E}) \tag{10}$$

$$\mathtt{S}^\sharp_{i+1} = \arg\min_{\mathtt{S}^\sharp} \mathcal{L}(\mathtt{S}, \mathtt{S}^\sharp_i, \mathtt{C}, \mathtt{E}) \tag{11}$$

$$\mathtt{E}_{i+1} = \arg\min_{\mathtt{E}} \mathcal{L}(\mathtt{S}, \mathtt{S}^\sharp, \mathtt{C}, \mathtt{E}_i) \tag{12}$$

$$\mathtt{C}_{i+1} = \arg\min_{\mathtt{C}} \mathcal{L}(\mathtt{S}, \mathtt{S}^\sharp, \mathtt{C}_i, \mathtt{E}) \tag{13}$$

Algorithm 1 Multi-body non-rigid structure-from-motion and segmentation via the ADMM

Require:
2D feature track matrix $\mathtt{W}$, camera motion $\mathtt{R}$, λ_1, λ_2, $\rho > 1$, β_m, ϵ;

Initialize: $\mathtt{S}^{(0)}$, $\mathtt{S}^{\sharp(0)}$, $\mathtt{C}^{(0)}$, $\mathtt{E}^{(0)}$, $\{\mathbf{Y}_i^{(0)}\}_{i=1}^4 = \mathbf{0}$, $\beta^{(0)}$;

while not converged **do**
1. Update $(\mathtt{S}, \mathtt{S}^\sharp, \mathtt{E}, \mathtt{C})$ by Eq. (10), Eq. (11), Eq. (12), Eq. (13) and Eq. (14);
2. Update $\{\mathbf{Y}_i\}_{i=1}^4$ and β by Eq. (15)-Eq. (17);
3. Check the convergence conditions $\|\mathtt{S}^\sharp - g(\mathtt{S})\|_\infty \leq \epsilon$, $\|\mathtt{S} - \mathtt{SC}\|_\infty \leq \epsilon$, $\|1^T\mathtt{C} - 1^T\|_\infty \leq \epsilon$, and $\|\mathtt{CD} - \mathtt{E}\|_\infty \leq \epsilon$;

end while

Ensure: $\mathtt{C}, \mathtt{S}, \mathtt{S}^\sharp$.
Form an affinity matrix $\mathtt{A} = |\mathtt{C}| + |\mathtt{C}^T|$, then apply spectral clustering [14] to $\mathtt{A}$.

$$\mathtt{C}_{i+1} = \mathtt{C}_i - \mathrm{diag}(\mathtt{C}_i), \tag{14}$$

Finally, the Lagrange multipliers $\{\mathtt{Y}_i\}_{i=1}^4$ and β are updated as:

$$\mathtt{Y}_1 = \mathtt{Y}_1 + \beta(\mathtt{S}^\sharp - g(\mathtt{S})), \mathtt{Y}_2 = \mathtt{Y}_2 + \beta(\mathtt{S} - \mathtt{SC}), \tag{15}$$

$$\mathtt{Y}_3 = \mathtt{Y}_3 + \beta(\mathtt{CD} - \mathtt{E}), \mathtt{Y}_4 = \mathtt{Y}_4 + \beta(1^T\mathtt{C} - 1^T), \tag{16}$$

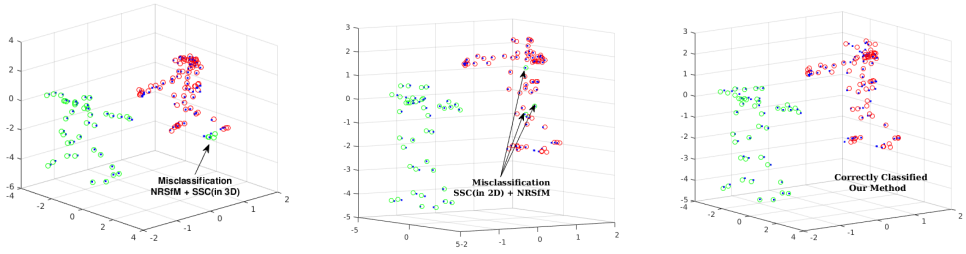
$$\beta = \min(\beta_m, \rho\beta), \tag{17}$$

4.1. Initialization

As our method tries to solve a non-convex optimization problem (9), a proper initialization is needed. In this paper, we initialize proper camera motion using Dai. *et al* approach [6]. In our current implementation, we have fixed the camera motion while updating the 3D non-rigid reconstruction and segmentation. In future, we will put the update of camera rotation in the loop. The initial structure $\mathtt{S}^{(0)}$ was initialized as pinv(R)* W. We kept $\beta^{(0)} = 1e^{-3}$ and $\rho = 1.1$ in all our experiments. *Note* : In the previous §4, we derive the solution when dense track features were provided as input. However, solution for sparse feature tracks case could be viewed as a simplified form of the dense case by removing the spatial constraint term $\|\mathtt{CD}\|$ directly.

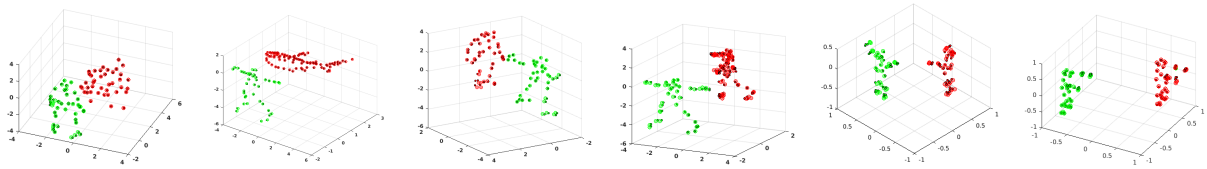
5. Experiments

To evaluate the effectiveness of our approach for multi-body non-rigid structure-from-motion, we conducted extensive experiments on both synthetic data and real images, under both sparse and dense scenarios.



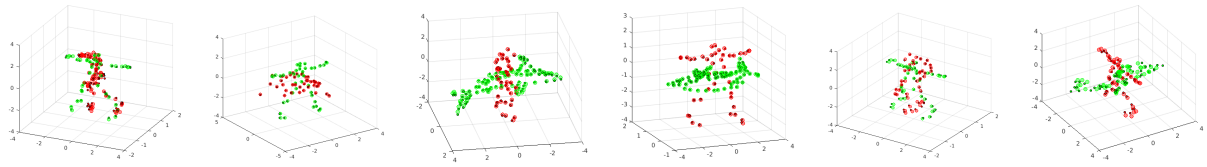
(a) (b) (c)

Figure 4: Demonstrating the efficacy of our approach. The above plot shows the results on Dance + Yoga sequence. (a) Result obtained by applying BMM method [6] to get 3D points and then use SSC [7] to segment 3D points. (b) Result obtained by applying SSC [7] to 2D feature tracks and then use BMM [6] separately to each segment to get 3D reconstruction. (c) Result by applying simultaneous reconstruction and segmentation framework (Our approach).



(a) (b) (c) (d) (e) (f)

Figure 5: 3D reconstruction and segmentation of different multi-body non-rigid motion sequences a) Face-Pickup Sequence; b) Shark-Yoga Sequence; c) Stretch-Yoga Sequence; d) Dance-Yoga Sequence; e) p3_ball_1; f) p4_meet_12. (a)-(d) CMU MoCap dataset [2], (e)-(f) UMPM dataset [20]. Dark small circles in the respective segments shows the Ground-Truth 3D points.



(a) (b) (c) (d) (e) (f)

Figure 6: 3D reconstruction and segmentation of different multi-body non-rigid motion sequences, when different objects intersect each other. a) Dance-Yoga; b) Face-Yoga Sequence; c) Shark-Stretch Sequence; d) Shark-Yoga Sequence; e) Stretch-Yoga Sequence; f) Walking-Yoga. (a)-(f) CMU MoCap dataset [2]. Dark small circles in the respective segments shows the Ground-Truth 3D points.

Table 2: Performance comparison between our method and the baseline methods, where 3D reconstruction error (e_{3D}) and non-rigid motion segmentation error (e_{MS}) are used as error metrics.

Dataset	BMM + SSC (3D)		SSC(2D) + BMM		Our Method	
	e_{3D}	e_{MS}	e_{3D}	e_{MS}	e_{3D}	e_{MS}
Dance + Yoga	0.0456	0.0345	0.0588	0.0259	0.046	**0.0**
Drink + Walking	0.0745	0.0	0.0858	0.0	**0.073**	**0.0**
Shark + Stretch	0.0246	0.4015	0.0979	0.3939	0.025	**0.0**
Walking + Yoga	0.0702	0.0	0.0900	0.0	**0.0702**	**0.0**
Face + Pickup	0.0324	0.0988	0.0239	0.0988	0.025	**0.0**
Face + Yoga	0.0172	0.012	0.0332	0.012	0.019	**0.0**
Shark + Yoga	0.0356	0.4167	0.1049	0.4091	0.0371	**0.0**
Stretch + Yoga	0.0392	0.0	0.0557	0.0	0.0393	**0.0**

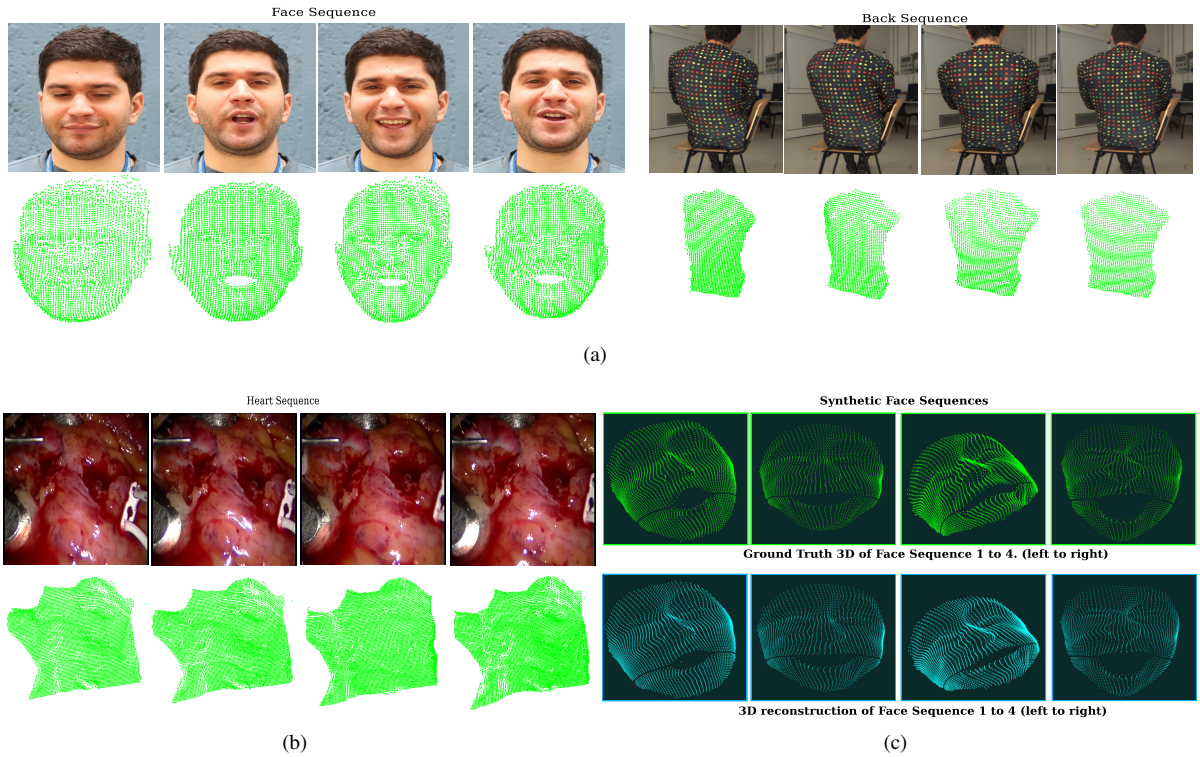


Figure 7: Reconstruction results on the real and synthetic data-sets: (a) - (b) Real Face, Back and Heart (c) Synthetic Face Sequence. [9].

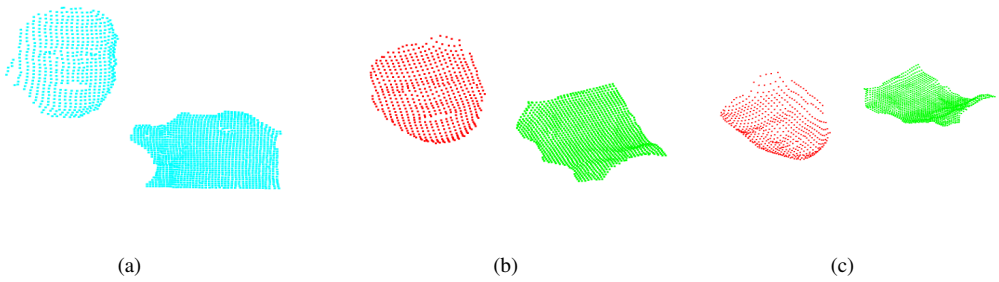


Figure 8: Experimental results on synthetic dense "Face+Heart" sequence. (a) Input 2D tracks of Face+Heart Sequence. (b)-(c) segmentation and reconstruction on the dense real sequence data-sets: (Face + Heart) Sequence [9]. Different color signature symbolizes the corresponding class labels. (Best Viewed on Screen)

As our method jointly perform non-rigid 3D reconstruction and segmentation, we use the following criteria to measure the performances of the algorithm:
(i) Relative error in multi-body non-rigid 3D reconstruction

$$e_{3D} = \|\mathtt{S}_f^{est.} - \mathtt{S}_f^{GT}\|_F / \|\mathtt{S}_f^{GT}\|_F, \tag{18}$$

(ii) Error in multi-body non-rigid motion segmentation,

$$e_{MS} = \frac{\text{total number of incorrectly segmented trajectories}}{\text{total number oftrajectories}}. \tag{19}$$

5.1. Multi-body non-rigid data

To advocate the performance of our framework, we tested it on both synthetic and real datasets. We synthesized multiple non-rigid objects by using the CMU Mocap dataset [2] and the UMPM dataset [20]. In Fig. 5, we illustrate six examples of multi-body NRSFM sequences and its results.

5.2. Performance comparison on sparse NRSFM

In Table 3, we compared the segmentation results of our approach with SSC [7] and EDSC [12] on multi-body non-

Table 3: Motion segmentation performance comparison with SSC [7] and EDSC [12] over 2D feature tracks.

Dataset	SSC (e_{MS})	EDSC (e_{MS})	Ours
Dance+Yoga	0.025	0.0345	0.0
Drink+Walking	0.0	0.01	0.0
Shark+Stretch	0.3939	0.0	0.0
Walking+Yoga	0.0	0.0	0.0
Face+Pickup	0.098	0.0	0.0
Face+Yoga	0.012	0.0	0.0
Shark+Yoga	0.41	0.0	0.0
Stretch+Yoga	0.0	0.0	0.0

rigid sequences over 2D feature tracks. It clearly demonstrates that performing non-rigid motion segmentation in 3D space using our approach leads to remarkable results. Since, our method jointly solves for 3D reconstruction and multi-body non-rigid motion segmentation, we compare our method with the two stage methods, namely

1) Baseline method 1: Single body NRSFM method (State-of-the-art "block-matrix method" [6] was used) followed by subspace clustering of the 3D trajectories (SSC [7] was used), denoted as "BMM+SSC(3D)";

2) Baseline method 2: Subspace clustering of the 2D feature tracks (2D trajectories) followed by single body NRSFM for each cluster of 2D feature tracks, denoted as "SSC(2D)+BMM".

Table 2 provides experimental comparisons between our method and the two baseline methods in dealing with multi-body NRSFM. In all the sequences, our method achieves zero multi-body non-rigid motion segmentation error and comparable 3D non-rigid reconstruction performance.

Experimental results show that our 3D reconstruction is very close to the accuracy of BMM [6] on almost all datasets. However, the advantage of our framework is that we can achieve robust segmentation along with better 3D reconstruction at the same time. Fig.4 shows the robustness of our approach. Our method faithfully reconstructs and segments two different complex non-rigid motions. Extensive experiments were performed on synthetic sparse data-sets with different combination of non-rigid motion Fig. 5(a), 5(b), 5(c) and 5(d) show some of the results on these different combinations of non-rigid motion on the CMU Mocap dataset [2], where Fig. 5(e) and 5(f) present results on the UMPM dataset [20]. Furthermore, we tested our approach in scenarios where different non-rigid moving objects intersect each other as shown in Figure 6(a)-6(f). Our method is able to reconstruct and segment each intersecting object.

Table 4: Quantitative results on synthetic face sequence, without and with neighboring constraints.

Dataset	#Features	e3d(only C)	e3d(with CD)
Face Seq. 1	3275	0.0749	0.0745
Face Seq. 2	3275	0.0506	0.050
Face Seq. 3	3275	0.0384	0.0380
Face Seq. 4	3275	0.0446	0.0443

5.3. Analysis on dense NRSFM

To expeditiously evaluate the effectiveness of our implementation over available dense sequences, we tested our method on uniformly sampled version of the original sequences for the sake of efficient implementation. We performed experimentation on benchmark NRSFM synthetic and real data-set sequence [9] introduced by Garg *et al.*. Table 4 reports the obtained 3D reconstruction error on these synthetic sequences [9]. Figure 7(a)-7(c) provide visual insight of the resultant 3D shapes over real and synthetic sequences. To test the segmentation of different structures on real dense dataset, we performed experiments by combining two real dataset sequence (Face+Heart). Figure 8(b) and 8(c) show the segmentation and reconstruction of dense non-rigid feature tracks to their corresponding classes. *Note: For more analysis and experiment results, please refer to supplementary material.*

6. Conclusions

This paper has filled in a missing gap in the structure-from-Motion family by proposing a new framework for multi-body non-rigid-structure-from-motion. It achieves a joint non-rigid reconstruction and non-rigid shape segmentation of multiple deformable structures observed in a single image sequence. Under our new multi-body NRSFM framework, the solutions for motion segmentation and the solutions for 3D reconstruction can better constrain each other. We achieved superior performance in both 3D non-rigid reconstruction and non-rigid motion segmentation, compared with the alternative, two stage methods (first segment, then reconstruct or first reconstruct, then segment). In future, we plan to investigate the scalability issue with our current implementation and apply the new method to more dense feature tracks in longer video sequences.

Acknowledgment

This work was supported in part by Australian Research Council (ARC) grants (DE140100180, DP120103896, LP100100588, CE140100016), Australia ARC Centre of Excellence Program on Roboitic Vision, NICTA (Data61) and Natural Science Foundation of China (61420106007).

References

[1] I. Akhter, Y. Sheikh, and S. Khan. In defense of orthonormality constraints for nonrigid structure from motion. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 1534–1541, 2009. 2

[2] I. Akhter, Y. Sheikh, S. Khan, and T. Kanade. Nonrigid structure from motion in trajectory space. In *Advances in Neural Information Processing Systems*, pages 41–48, 2008. 1, 2, 6, 7, 8

[3] C. Bregler, A. Hertzmann, and H. Biermann. Recovering non-rigid 3D shape from image streams. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 690–696, 2000. 1, 2, 3

[4] J. Cho, M. Lee, and S. Oh. Complex non-rigid 3d shape recovery using a procrustean normal distribution mixture model. *International Journal of Computer Vision*, pages 1–21, 2015. 1, 2

[5] J. P. Costeira and T. Kanade. A multibody factorization method for independently moving objects. *International Journal of Computer Vision*, 29(3):159–179. 1, 3

[6] Y. Dai, H. Li, and M. He. A simple prior-free method for non-rigid structure-from-motion factorization. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 2018–2025, 2012. 1, 2, 4, 5, 6, 8

[7] E. Elhamifar and R. Vidal. Sparse subspace clustering: Algorithm, theory, and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(11):2765–2781, 2013. 2, 4, 6, 7, 8

[8] K. Fragkiadaki, M. Salas, P. A. Arbeláez, and J. Malik. Grouping-based low-rank trajectory completion and 3d reconstruction. In *Advances in Neural Information Processing Systems*, pages 55–63, 2014. 2

[9] R. Garg, A. Roussos, and L. Agapito. Dense variational reconstruction of non-rigid surfaces from monocular video. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 1272–1279, 2013. 1, 2, 4, 7, 8

[10] R. Garg, A. Roussos, and L. Agapito. A variational approach to video registration with subspace constraints. *International Journal of Computer Vision*, 104(3):286–314, 2013. 2

[11] P. Gotardo and A. Martinez. Non-rigid structure from motion with complementary rank-3 spaces. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 3065–3072, 2011. 2

[12] P. Ji, M. Salzmann, and H. Li. Efficient dense subspace clustering. In *IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 461–468, March 2014. 7, 8

[13] M. Lee, J. Cho, C.-H. Choi, and S. Oh. Procrustean normal distribution for non-rigid structure from motion. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 1280–1287, 2013. 1, 2

[14] A. Y. Ng, M. I. Jordan, and Y. Weiss. On spectral clustering: Analysis and an algorithm. In T. G. Dietterich, S. Becker, and Z. Ghahramani, editors, *Advances in Neural Information Processing Systems 14*, pages 849–856. MIT Press, 2002. 4, 5

[15] R. Ranftl, V. Vineet, Q. Chen, and V. Koltun. Dense monocular depth estimation in complex dynamic scenes. In *Proc. IEEE Conf. Comp. Vis. Patt. Recogn.*, June 2016. 3

[16] C. Russell, J. Fayad, and L. Agapito. Dense non-rigid structure from motion. In *International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT)*, pages 509–516, 2012. 2

[17] C. Russell, R. Yu, and L. Agapito. *Computer Vision – ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part VII*, chapter Video Pop-up: Monocular 3D Reconstruction of Dynamic Scenes, pages 583–598. Springer International Publishing, Cham, 2014. 2

[18] T. Simon, J. Valmadre, I. Matthews, and Y. Sheikh. Separable spatiotemporal priors for convex reconstruction of time-varying 3d point clouds. In *European Conference on Computer Vision*, pages 204–219. 2014. 2

[19] C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: a factorization method. *Int'l J. Computer Vision*, 9(2):137–154, 1992. 3

[20] N. van der Aa, X. Luo, G. Giezeman, R. Tan, and R. Veltkamp. Umpm benchmark: A multi-person dataset with synchronized video and motion capture data for evaluation of articulated human motion and interaction. In *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*, pages 1264–1269, Nov 2011. 6, 7, 8

[21] J. Xiao, J. Chai, and T. Kanade. A closed-form solution to non-rigid shape and motion recovery. In *Proc. European Conf. Computer Vision*, volume 3024, pages 573–587, 2004. 1, 2

[22] J. Yan and M. Pollefeys. A factorization-based approach for articulated nonrigid shape, motion and kinematic chain recovery from video. *IEEE Trans. Pattern Anal. Mach. Intell.*, 30(5):865–877, 2008. 1, 3

[23] R. Yu, C. Russell, N. D. F. Campbell, and L. Agapito. Direct, dense, and deformable: Template-based non-rigid 3d reconstruction from rgb video. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 918–926, Dec 2015. 2

[24] Y. Zhu, D. Huang, F. De La Torre, and S. Lucey. Complex non-rigid motion 3d reconstruction by union of subspaces. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 1542–1549, June 2014. 2

[25] Y. Zhu and S. Lucey. Convolutional sparse coding for trajectory reconstruction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(3):529–540, March 2015. 2

Multi-body Non-rigid Structure-from-Motion Supplementary Material

Suryansh Kumar[1] Yuchao Dai[1] Hongdong Li[1,2]
[1]Research School of Engineering, Australian National University.
[2] Australian Center for Robotic Vision.

Abstract

In the supplementary material, we provide a detailed derivation of each sub-problems in the formulation. Besides this, we also provide insight into convergence curve and effect of noisy track features on the performance of our algorithm.

1. Sub-problem derivation of the involved optimization

$$\begin{aligned}&\underset{\mathtt{C},\mathtt{S},\mathtt{S}^\sharp}{\text{minimize}}\ \frac{1}{2}\|\mathtt{W}-\mathtt{RS}\|_F^2+\lambda_1\|\mathtt{C}\|_1+\lambda_2\|\mathtt{CD}\|_1+\lambda_3\|\mathtt{S}^\sharp\|_*\\&\text{subject to:}\\&\mathtt{S}^\sharp=g(\mathtt{S}),\mathtt{S}=\mathtt{SC},1^T\mathtt{C}=1^T,\text{diag}(\mathtt{C})=0.\end{aligned}\tag{1}$$

To further decouple the constraint, we introduce an auxiliary variable $\mathtt{E}=\mathtt{CD}$. With these operations, the optimization problem Eq.-(1) can be reformulated as:

$$\begin{aligned}&\underset{\mathtt{E},\mathtt{S},\mathtt{S}^\sharp,\mathtt{C}}{\text{minimize}}\ \frac{1}{2}\|\mathtt{W}-\mathtt{RS}\|_F^2+\lambda_1\|\mathtt{E}\|_1+\lambda_2\|\mathtt{S}^\sharp\|_*\\&\text{subject to:}\\&\mathtt{S}^\sharp=g(\mathtt{S}),\mathtt{S}=\mathtt{SC},\mathtt{CD}=\mathtt{E},1^T\mathtt{C}=1^T,\text{diag}(\mathtt{C})=0.\end{aligned}\tag{2}$$

The Augmented Lagrangian formulation for Eq.-(2) is:

$$\begin{aligned}\mathcal{L}(\mathtt{S},\mathtt{S}^\sharp,\mathtt{C},\mathtt{E},\{\mathtt{Y}_i\}_{i=1}^4)=\frac{1}{2}\|\mathtt{W}-\mathtt{RS}\|_F^2+\lambda_1\|\mathtt{E}\|_1+\\\lambda_2\|\mathtt{S}^\sharp\|_*+<\mathtt{Y}_1,\mathtt{S}^\sharp-g(\mathtt{S})>+\frac{\beta}{2}\|\mathtt{S}^\sharp-g(\mathtt{S})\|_F^2+\\<\mathtt{Y}_2,\mathtt{S}-\mathtt{SC}>+\frac{\beta}{2}\|\mathtt{S}-\mathtt{SC}\|_F^2+<\mathtt{Y}_3,\mathtt{CD}-\mathtt{E}>+\\\frac{\beta}{2}\|\mathtt{CD}-\mathtt{E}\|_F^2+<\mathtt{Y}_4,1^T\mathtt{C}-1^T>+\frac{\beta}{2}\|1^T\mathtt{C}-1^T\|_F^2.\end{aligned}$$

1.0.1 The solution of S:

$$\begin{aligned}\mathtt{S}=\arg\min_{\mathtt{S}}\ &\tfrac{1}{2}\|\mathtt{W}-\mathtt{RS}\|_F^2+<\mathtt{Y}_1,\mathtt{S}^\sharp-g(\mathtt{S})>+\\&\tfrac{\beta}{2}\|\mathtt{S}^\sharp-g(\mathtt{S})\|_F^2+<\mathtt{Y}_2,\mathtt{S}-\mathtt{SC}>+\\&\tfrac{\beta}{2}\|\mathtt{S}-\mathtt{SC}\|_F^2.\end{aligned}\tag{3}$$

The sub-problem for S reaches a least squares problem. The closed-form solution of S can be derived as:

$$\begin{aligned}&\tfrac{1}{\beta}(\mathtt{R}^T\mathtt{R}+\beta\mathtt{I})\mathtt{S}+\mathtt{S}(\mathtt{I}-\mathtt{C})(\mathtt{I}-\mathtt{C}^T)=\tfrac{1}{\beta}\mathtt{R}^T\mathtt{W}+\\&(g^{-1}(\mathtt{S}^\sharp)+\tfrac{g^{-1}(\mathtt{Y}_1)}{\beta}-\tfrac{\mathtt{Y}_2}{\beta}(\mathtt{I}-\mathtt{C}^T)),\end{aligned}\tag{4}$$

which is a Sylvester equation.

1.0.2 The solution of $\mathtt{S}^\sharp$:

$$\mathtt{S}^\sharp=\arg\min_{\mathtt{S}^\sharp}\ \lambda_2\|\mathtt{S}^\sharp\|_*+<\mathtt{Y}_1,\mathtt{S}^\sharp-g(\mathtt{S})>+\frac{\beta}{2}\|\mathtt{S}^\sharp-g(\mathtt{S})\|_F^2\tag{5}$$

A close-form solution exists for this sub-problem. Let's define the soft-thresholding operation as $\mathcal{S}_\tau[x]=\text{sign}(x)\max(|x|-\tau,0)$. The optimal solution to Eq.-(5) can be obtained as:

$$\mathtt{S}^\sharp=\mathtt{U}\mathcal{S}_{\lambda_2/\beta}(\Sigma)\mathtt{V},\tag{6}$$

where $[\mathtt{U},\Sigma,\mathtt{V}]=\text{svd}(g(\mathtt{S})-\mathtt{Y}_1/\beta)$.

1.0.3 The solution of E:

$$\mathtt{E}=\arg\min_{\mathtt{E}}\ \lambda_1\|\mathtt{E}\|_1+<\mathtt{Y}_3,\mathtt{CD}-\mathtt{E}>+\frac{\beta}{2}\|\mathtt{CD}-\mathtt{E}\|_F^2,\tag{7}$$

A close-form solution exists for this sub-problem by using element-wise shrinkage.

$$\mathtt{E}=\mathcal{S}_{\lambda_1/\beta}(\mathtt{CD}+\frac{\mathtt{Y}_3}{\beta}).\tag{8}$$

1.0.4 The solution of C:

$$\begin{aligned}\mathtt{C}=\arg\min_{\mathtt{C}}\ &<\mathtt{Y}_2,\mathtt{S}-\mathtt{SC}>+\tfrac{\beta}{2}\|\mathtt{S}-\mathtt{SC}\|_F^2+\\&<\mathtt{Y}_3,\mathtt{CD}-\mathtt{E}>+\tfrac{\beta}{2}\|\mathtt{CD}-\mathtt{E}\|_F^2+<\mathtt{Y}_4,1^T\mathtt{C}-1^T>+\\&\tfrac{\beta}{2}\|1^T\mathtt{C}-1^T\|_F^2.\end{aligned}\tag{9}$$

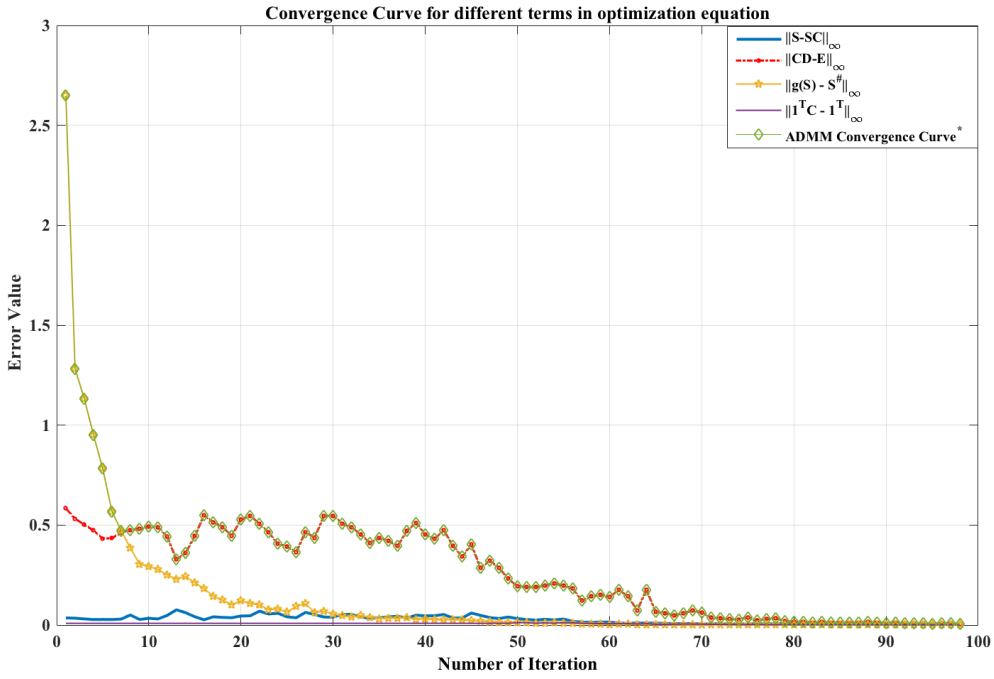


Figure 1: Typical convergence curves of the objective function and the primal residuals $\|\mathtt{S}^\sharp - g(\mathtt{S})\|_\infty, \|\mathtt{S} - \mathtt{SC}\|_\infty, \|\mathtt{CD} - \mathtt{E}\|_\infty$ and $\|1^T\mathtt{C} - 1^T\|_\infty$. The above plot shows the convergence statistics for Dance+Yoga Sequence. *ADMM Convergence Curve = maximum of ($\|\mathtt{S}^\sharp - g(\mathtt{S})\|_\infty, \|\mathtt{S} - \mathtt{SC}\|_\infty, \|\mathtt{CD} - \mathtt{E}\|_\infty$ and $\|1^T\mathtt{C} - 1^T\|_\infty$).

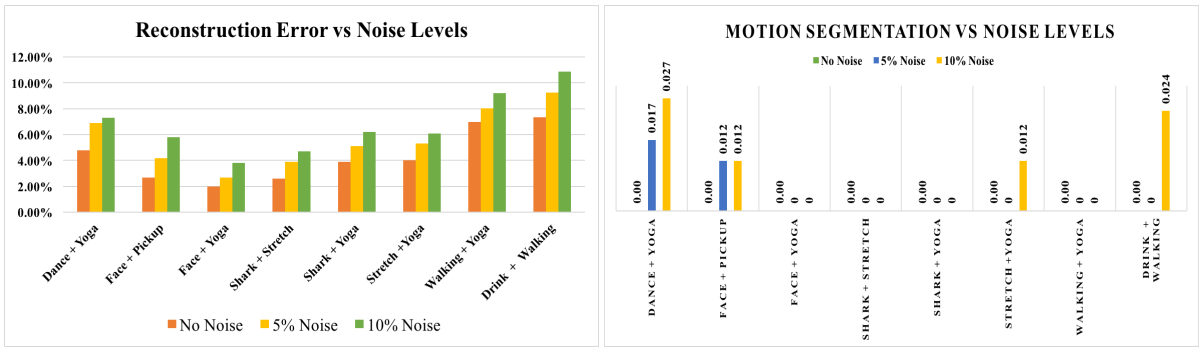


Figure 2: Left: 3D Reconstruction error VS noise levels; Right: non-rigid motion segmentation error VS noise levels.

The closed-form solution of $\mathtt{C}$ is derived as:

$$(\mathtt{S}^T\mathtt{S} + 1\mathtt{1}^T)\mathtt{C} + \mathtt{C}(\mathtt{D}\mathtt{D}^T) = \mathtt{S}^T\mathtt{S} + \mathtt{S}^T\frac{\mathtt{Y}_2}{\beta} + \mathtt{E}\mathtt{D}^T - \mathtt{Y}_3\frac{\mathtt{D}^T}{\beta} + 1\mathtt{1}^T - 1\frac{\mathtt{Y}_4}{\beta}. \tag{10}$$

$$\mathtt{C} = \mathtt{C} - \mathrm{diag}(\mathtt{C}), \tag{11}$$

Finally, the Lagrange multipliers $\{\mathtt{Y}_i\}_{i=1}^4$ and β are updated as:

$$\mathtt{Y}_1 = \mathtt{Y}_1 + \beta(\mathtt{S}^\sharp - g(\mathtt{S})), \mathtt{Y}_2 = \mathtt{Y}_2 + \beta(\mathtt{S} - \mathtt{SC}), \tag{12}$$

$$\mathtt{Y}_3 = \mathtt{Y}_3 + \beta(\mathtt{CD} - \mathtt{E}), \mathtt{Y}_4 = \mathtt{Y}_4 + \beta(1^T\mathtt{C} - 1^T), \tag{13}$$

$$\beta = \min(\beta_m, \rho\beta), \tag{14}$$

1.1. Experiment: Convergence

In this experiment, we would like to study the convergence of our algorithm. Given noise free input, we want to check whether or not our proposed algorithm converge; and if it does converge, whether it converges to the correct

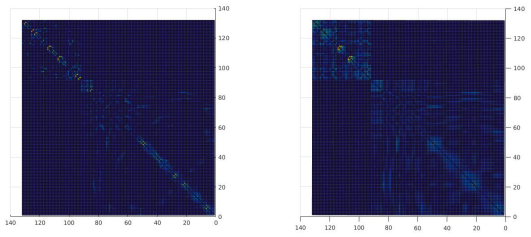


Figure 3: Obtained Affinity Matrix $\mathtt{A} = |\mathtt{C}| + |\mathtt{C}^T|$. a) Affinity matrix from SSC; b) Affinity matrix from our Method. Best Viewed on Screen.

solution. Note that we use the sparse sequences from the CMU MoCap dataset [1] directly without any dimension reduction or projection. Typical convergence curves of the objective function and the primal residuals are illustrated in Fig. 1.

1.2. Experiment: Performance on noisy feature tracks

In the second experiment, we conducted analysis to the performance of our method under different level of noise. In the same manner as above, we generated multi-body non-rigid sequences ("Dance + Yoga", "Face + Pickup", "Face + Yoga", "Shark + Stretch", "Shark + Yoga", "Stretch + Yoga" and "Walking + Yoga"), then zero-mean Gaussian noise with standard deviation σ were added to the feature tracks. For each noisy input, we ran our code for 5 times and recorded the mean 3D reconstruction error and non-rigid motion segmentation error.

In Fig. 2, we illustrated the statistical results of 3D non-rigid reconstruction and non-rigid motion segmentation. From the figures, we conclude that both the 3D reconstruction error and the motion segmentation error increases with the increase of noise level. Our 3D reconstruction based non-rigid motion segmentation achieves smaller motion segmentation error compared with 2D trajectory based motion segmentation methods such as sparse subspace clustering (SSC) [2] and efficient dense subspace clustering (EDSC) [3].

1.3. Affinity Matrix Comparison

In Fig. 3, we compare the affinity matrices from SSC [2] and our method. It is clear that our method outputs an affinity matrix with better structure, which results in better non-rigid motion segmentation performance.

References

[1] I. Akhter, Y. Sheikh, S. Khan, and T. Kanade. Nonrigid structure from motion in trajectory space. In *Advances in Neural Information Processing Systems*, pages 41–48, 2008. 3

[2] E. Elhamifar and R. Vidal. Sparse subspace clustering: Algorithm, theory, and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(11):2765–2781, 2013. 3

[3] P. Ji, M. Salzmann, and H. Li. Efficient dense subspace clustering. In *IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 461–468, March 2014. 3